

What is claimed is:

- 1     1.     A method of recovering a stripe of erasure coded data stored across a plurality  
2         of storage devices comprising the steps of:  
3             sending query messages which include a timestamp indicating a  
4             current time to the storage devices;  
5             receiving query reply messages from at least a first quorum of the  
6             storage devices including a minimum number of stripe blocks needed to  
7             decode the stripe of erasure coded data, the minimum number of the stripe  
8             blocks having matching validation timestamps;  
9             encoding the stripe of erasure coded data from the minimum number of  
10            the stripe blocks;  
11            sending a write message to each of the storage devices including the  
12            timestamp and the stripe block destined for the storage device; and  
13            receiving a write reply message from each of at least a second quorum  
14            of the storage devices indicating that the stripe block was successfully  
15            stored, the first and second quorums each meeting a quorum condition of a  
16            number such that any two selections of the number of the stripe blocks  
17            intersect in the minimum number of the stripe blocks.
- 1     2.     The method of claim 1 further comprising the step of setting a max timestamp  
2         to an initial high value.
- 1     3.     The method of claim 2 wherein each of the query messages include the max  
2         timestamp.
- 1     4.     The method of claim 3 wherein the query reply message from each of the first  
2         quorum of the storage devices indicate that the timestamp is later than a pending  
3         write timestamp.
- 1     5.     The method of claim 4 wherein the query reply message from each of the first  
2         quorum of the storage devices indicate that the timestamp is later than a validation  
3         timestamp for a latest version of the stripe block stored in the storage device.

- 1    6.     The method of claim 5 wherein each of the query reply messages include a  
2        version of the stripe block having the validation timestamp earlier than the max  
3        timestamp.
- 1    7.     The method of claim 6 wherein each of the query reply messages include the  
2        validation timestamp corresponding to the version of the stripe block.
- 1    8.     The method of claim 7 further comprising the step of setting the max  
2        timestamp to the validation timestamp within the query reply messages having a  
3        latest time.
- 1    9.     The method of claim 8 further comprising the step of identifying the stripe  
2        blocks having the validation timestamp of the max timestamp as max timestamp  
3        stripe blocks.
- 1    10.    The method of claim 9 further comprising the step of decoding the stripe of  
2        erasure coded data from the max timestamp stripe blocks if a quantity of the max  
3        timestamp stripe blocks is at least the minimum number.
- 1    11.    The method of claim 10 further comprising the step of returning to the step of  
2        sending the query messages if the quantity of the max timestamp stripe blocks is  
3        less than the minimum number.
- 1    12.    The method of claim 10 wherein each of the query reply messages from the  
2        second quorum of the storage devices indicate that the timestamp is later than a  
3        validation timestamp.
- 1    13.    The method of claim 12 wherein each of the query reply messages from the  
2        second quorum of the storage devices indicate that the timestamp is no earlier than  
3        the pending write timestamp.
- 1    14.    A method of recovering a stripe of erasure coded data comprising the steps of:  
2        generating a timestamp;  
3        setting a max timestamp to an initial high value;

4            sending a query message to each of a plurality of storage devices upon  
5            which the stripe of erasure coded data is stored, the stripe of erasure coded  
6            data comprising stripe blocks, each of the query messages including the  
7            timestamp and the max timestamp;

8            receiving a query reply message from each of at least a first quorum of  
9            the storage devices indicating that the timestamp is later than a pending  
10           write timestamp and that the timestamp is later than a validation timestamp  
11           for a latest version of a stripe block stored in the storage device, each of  
12           the query reply messages including a version of the stripe block having the  
13           validation timestamp earlier than the max timestamp, each of the query  
14           reply messages including the validation timestamp corresponding to the  
15           version of the stripe block;

16           setting the max timestamp to the validation timestamp within the query  
17           reply messages having a latest time;

18           identifying the stripe blocks having the validation timestamp of the  
19           max timestamp as max timestamp stripe blocks;

20           decoding the stripe of erasure coded data from the max timestamp  
21           stripe blocks if a quantity of the max timestamp stripe blocks is at least a  
22           minimum number of the stripe blocks needed to decode the stripe of  
23           erasure coded data, thereby forming data blocks;

24           encoding the stripe of erasure coded data from the data blocks, thereby  
25           forming the stripe blocks;

26           sending a write message to each of the storage devices including the  
27           timestamp and the stripe block destined for the storage device; and

28           receiving a write reply message from each of at least a second quorum  
29           of the storage devices indicating that the timestamp is later than the  
30           validation timestamp, that the timestamp is no earlier than the pending  
31           write timestamp, and that the stripe block was successfully stored, the first  
32           and second quorums each meeting a quorum condition of a number such  
33           that any two selections of the number of the stripe blocks intersect in the  
34           minimum number of the stripe blocks.

- 1    15.    The method of claim 14 further comprising the step of returning to the step of  
2           sending the query messages if the quantity of the max timestamp stripe blocks is

3 less than the minimum number.

1 16. A computer readable memory comprising computer code for implementing a  
2 method of recovering a stripe of erasure coded data, the method of recovering the  
3 stripe of erasure coded data comprising the steps of:

4 sending query messages to the storage devices which include a  
5 timestamp indicating a current time;

6 receiving query reply messages from at least a first quorum of the  
7 storage devices including a minimum number of the stripe blocks needed  
8 to decode the stripe of erasure coded data, the minimum number of the  
9 stripe blocks having matching validation timestamps;

10 encoding the stripe of erasure coded data from the minimum number of  
11 the stripe blocks;

12 sending a write message to each of the storage devices including the  
13 timestamp and the stripe block destined for the storage device; and

14 receiving a write reply message from at least a second quorum of the  
15 storage devices indicating that the stripe block was successfully stored, the  
16 first and second quorums each meeting a quorum condition of a number  
17 such that any two selections of the number of the stripe blocks intersect in  
18 the minimum number of the stripe blocks.

1 17. The computer readable memory of claim 16 further comprising the step of  
2 setting a max timestamp to an initial high value.

1 18. The computer readable memory of claim 17 wherein each of the query  
2 messages include the max timestamp.

1 19. The computer readable memory of claim 18 wherein each of the query reply  
2 messages from the first quorum of the storage devices indicate that the timestamp  
3 is later than a pending write timestamp.

1 20. The computer readable memory of claim 19 wherein each of the query reply  
2 messages from the first quorum of the storage devices indicate that the timestamp  
3 is later than a validation timestamp for a latest version of the stripe block stored in

4       the storage device.

1   21.    The computer readable memory of claim 20 wherein each of the query reply  
2       messages include a version of the stripe block having the validation timestamp  
3       earlier than the max timestamp.

1   22.    The computer readable memory of claim 21 wherein each of the query reply  
2       messages include the validation timestamp corresponding to the version of the  
3       stripe block.

1   23.    The computer readable memory of claim 22 further comprising the step of  
2       setting the max timestamp to the validation timestamp within the query reply  
3       messages having a latest time.

1   24.    The computer readable memory of claim 23 further comprising the step of  
2       identifying the stripe blocks having the validation timestamp of the max  
3       timestamp as max timestamp stripe blocks.

1   25.    The computer readable memory of claim 24 further comprising the step of  
2       decoding the stripe of erasure coded data from the max timestamp stripe blocks if  
3       a quantity of the max timestamp stripe blocks is at least the minimum number.

1   26.    The computer readable memory of claim 25 further comprising the step of  
2       returning to the step of sending the query messages if the quantity of the max  
3       timestamp stripe blocks is less than the minimum number.

1   27.    The computer readable memory of claim 25 wherein each of the query reply  
2       messages from the second quorum of the storage devices indicate that the  
3       timestamp is later than a validation timestamp.

1   28.    The computer readable memory of claim 27 wherein each of the query reply  
2       messages from the second quorum of the storage devices indicate that the  
3       timestamp is no earlier than the pending write timestamp.

1   29.    A computer readable memory comprising computer code for implementing a  
2        method of recovering a stripe of erasure coded data, the method of recovering the  
3        stripe of erasure coded data comprising the steps of:  
4            generating a timestamp;  
5            setting a max timestamp to an initial high value;  
6            sending a query message to each of a plurality of storage devices upon  
7        which the stripe of erasure coded data is stored, the stripe of erasure coded  
8        data comprising stripe blocks, each of the query messages including the  
9        timestamp and the max timestamp;  
10          receiving a query reply message from each of at least a first quorum of  
11        the storage devices indicating that the timestamp is later than a pending  
12        write timestamp and that the timestamp is later than a validation timestamp  
13        for a latest version of a stripe block stored in the storage device, each of  
14        the query reply messages including a version of the stripe block having the  
15        validation timestamp earlier than the max timestamp, each of the query  
16        reply messages including the validation timestamp corresponding to the  
17        version of the stripe block;  
18          setting the max timestamp to the validation timestamp within the query  
19        reply messages having a latest time;  
20          identifying the stripe blocks having the validation timestamp of the  
21        max timestamp as max timestamp stripe blocks;  
22          decoding the stripe of erasure coded data from the max timestamp  
23        stripe blocks if a quantity of the max timestamp stripe blocks is at least a  
24        minimum number of the stripe blocks needed to decode the stripe of  
25        erasure coded data, thereby forming data blocks;  
26          encoding the stripe of erasure coded data from the data blocks, thereby  
27        forming the stripe blocks;  
28          sending a write message to each of the storage devices including the  
29        timestamp and the stripe block destined for the storage device; and  
30          receiving a write reply message from each of at least a second quorum  
31        of the storage devices indicating that the timestamp is later than the  
32        validation timestamp, that the timestamp is no earlier than the pending  
33        write timestamp, and that the stripe block was successfully stored, the first  
34        and second quorums each meeting a quorum condition of a number such

35                   that any two selections of the number of the stripe blocks intersect in the  
36                   minimum number of the stripe blocks.

1    30.    The computer readable memory of claim 29 further comprising the step of  
2           returning to the step of sending the query messages if the quantity of the max  
3           timestamp stripe blocks is less than the minimum number.